YAEN
09/830,647
SEQ ID: 1-4

GenCore version 4.5

OM protein - protein search, using sw model

Run on: December 27, 2001, 16:44:40 ; Search time 41.26 Seconds
(without alignments)
1210.019 Million cell updates/sec

Title: US-09-830-647-1
Perfect score: 3510
Sequence: 1 MNSGAMRIHSKGHFQGGIQV.....SDNLLTAFFSSPSTSTFTGF 674

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_1101:*
1: /SIDS2/gcgdata/geneseq/geneseqp/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqp/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqp/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqp/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqp/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqp/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqp/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqp/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqp/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqp/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqp/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqp/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqp/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqp/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqp/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqp/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqp/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqp/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqp/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqp/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqp/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqp/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	3510	100.0	674	21	AAB42482	Human ORFX ORF2246
2	3510	100.0	674	21	AAB03758	Human H37 amino ac
3	1661	47.3	320	22	AAM17195	Peptide #3629 enco
4	1661	47.3	320	22	AAM04892	Peptide #3574 enco
5	1283	36.6	245	22	AAB95232	Human protein sequ
6	1170	33.3	234	21	AAB03759	Human H37 amino ac
7	308	8.8	60	22	AAM19849	Peptide #6283 enco
8	308	8.8	60	22	AAM33333	Peptide #7370 enco
9	281	8.0	55	22	AAM32811	Peptide #6848 enco
10	259	7.4	49	22	AAM18938	Peptide #5372 enco
11	259	7.4	49	22	AAM31512	Peptide #5549 enco
12	208	5.9	170	22	AAB95297	
13	170	4.8	1516	21	AAB18195	
14	164	4.7	646	21	AAB18188	
15	164	4.7	1087	20	AAY19935	
16	164	4.7	1119	20	AAY19934	
17	164	4.7	1279	22	AAM39101	
18	164	4.7	1305	22	AAM40887	
19	162	4.6	783	11	AAR05804	
20	159.5	4.5	1392	20	AAY06999	
21	159	4.5	2663	22	AAM39097	
22	159	4.5	2688	22	AAM40883	
23	158.5	4.5	1427	12	AAR10534	
24	158	4.5	980	21	AAB18294	
25	157.5	4.5	1979	21	AAB18171	
26	156.5	4.5	3248	17	AAR99795	
27	155.5	4.4	2482	16	AAR72826	Human mitosin. Ho
28	155.5	4.4	2482	19	AAW23996	Human mitosin amin
29	152.5	4.3	1308	22	AAU14791	Novel bone marrow
30	152.5	4.3	1501	22	AAU14725	Novel bone marrow
31	150.5	4.3	2781	21	AAY57453	Human transcriptio
32	150.5	4.3	2907	21	AAY57452	Human transcriptio
33	147.5	4.2	2954	20	AAY01632	Amino acid sequenc
34	146.5	4.2	1257	22	AAB66926	RBP1 protein. Hom
35	146.5	4.2	1863	17	AAR81509	BRCA1 mutant from
36	145.5	4.1	1639	19	AAW54145	P. falciparum synt
37	144	4.1	688	21	AAY57274	E. canis 120 kDa i
38	144	4.1	2391	15	AAR55694	Carbamoyl-phosphat
39	142.5	4.1	761	20	AAY28934	CEV14 gene product
40	142.5	4.1	777	22	AAM13804	Peptide #238 encod
41	142.5	4.1	777	22	AAM26203	Peptide #240 encod
42	142.5	4.1	777	22	AAM01549	Peptide #231 encod
43	141.5	4.0	1863	17	AAR81493	BRCA1 mutant from
44	141.5	4.0	1863	17	AAR81511	BRCA1 mutant from
45	141	4.0	1036	21	AAG31888	Arabidopsis thalia

ALIGNMENTS

RESULT 1
AAB42482
ID AAB42482 standard; Protein; 674 AA.
XX
AC AAB42482;
XX
DT 08-FEB-2001 (first entry)
XX
DE Human ORFX ORF2246 polypeptide sequence SEQ ID NO:4492.
XX
KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW vulnerary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW thrombosis; contraceptive.
XX
OS Homo sapiens.
XX
PN WO200058473-A2.
XX
PD 05-OCT-2000.
XX
PF 31-MAR-2000; 2000WO-US08621.
XX
PR 31-MAR-1999; 99US-0127607.

```
PR   02-APR-1999;   99US-0127636.
PR   05-APR-1999;   99US-0127728.
PR   30-MAR-2000; 2000US-0540763.
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Shimkets RA,  Leach M;
XX
DR   WPI; 2000-602362/57.
DR   N-PSDB; AAC76691.
XX
PT   Novel nucleic acids and peptides derived from open reading frame X,
PT   useful for treating e.g. cancers, proliferative disorders,
PT   neurodegenerative disorders and cardiovascular disease -
XX
PS   Claim 11; Page 3676-3677; 5507pp; English.
XX
CC   AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
CC   which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC   sequences have activities such as: cytostatic; hepatotropic; vulnerary;
CC   antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
CC   osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
CC   immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
CC   antidiabetic; hypotensive; dermatological; immunosuppressive;
CC   antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
CC   antithyroid; and antianaemic. The sequences can be used for determining
CC   the presence of or predisposition to, or preventing or treating
CC   pathological conditions associated with an ORFX-associated disorder. The
CC   nucleic acids can be used to express ORFX proteins in gene therapy
CC   vectors. The proteins and nucleic acids may be used to treat cancers,
CC   proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC   graft vs host disease, cardiovascular disease, diabetes mellitus,
CC   hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
CC   erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
CC   bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC   allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
CC   nocturnal haemoglobinuria, antiinflammatory disease; to enhance
CC   coagulation; to inhibit thrombosis; and as a contraceptive.
XX
SQ   Sequence   674 AA;

  Query Match            100.0%;  Score 3510;  DB 21;  Length 674;
  Best Local Similarity   100.0%;  Pred. No. 2.1e-257;
  Matches  674;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 MNSGAMRIHSKGHFQGGIQVKNEKNRPSLKSLKTDNRPEKSKCKPLWGKVFYLDLPSVTI 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 mnsgamrihskghfqggiqvkneknrpslksltkdnrpekskckplwgkvfyldlpsvti 60

Qy      61 SEKLQKDIKDLGGRVEEFLSKDISYLISNKKEAKFAQTLGRISPVPSPESAYTAETTSPH 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 seklqkdikdlggrveeflskdisylisnkkeakfaqtlgrispvpspesaytaettsph 120

Qy     121 PSHDGSSFKSPDTVCLSRGKLLVEKAIKDHDFIPSNSILSNALSWGVKILHIDDIRYYIE 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 pshdgssfkspdtvclsrgkllvekaikdhdfipsnsilsnalswgvkilhiddiryyie 180

Qy     181 QKKKELYLLKKSSTSVRDGGKRVGSGAQKTRTGRLKKPFVKVEDMSQLYRPFYLQLTNMP 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 qkkkelyllkksstsvrdggkrvgsgaqktrtgrlkkpfvkvedmsqlyrpfylqltnmp 240

Qy     241 FINYSIQKPCSPFDVDKPSSMQKQTQVKLRIQTDGDKYGGTSIQLQLKEKKKKGYCECCL 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 finysiqkpcspfdvdkpssmqkqtqvklriqtdgdkyggtsiqlqlkekkkkgyceccl 300

Qy     301 QKYEDLETHLLSEQHRNFAQSNQYQVVDDIVSKLVFDFVEYEKDTPKKKRIKYSVGSLSP 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 qkyedlethllseqhrnfaqsnqyqvvddivsklvfdfveyekdtpkkkrikysvgslsp 360

Qy     361 VSASVLKKTEQKEKVELQHISQKDCQEDDTTVKEQNFLYKETQETEKKLLFISEPIPHPS 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 vsasvlkkteqkekvelqhisqkdcqeddttvkeqnflyketqetekkllfisepiphps 420

Qy     421 NELRGLNEKMSNKCSMLSTAEDDIRQNFTQLPLHKNKQECILDISEHTLSENDLEELRVD 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 nelrglnekmsnkcsmlstaeddirqnftqlplhknkqecildisehtlsendleelrvd 480

Qy     481 HYKCNIQASVHVSDFSTDNSGSQPKQKSDTVLFPAKDLKEKDLHSIFTHDSGLITINSSQ 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 hykcniqasvhvsdfstdnsgsqpkqksdtvlfpakdlkekdlhsifthdsglitinssq 540

Qy     541 EHLTVQAKAPFHTPPEEPNECDFKNMDSLPSGKIHRKVKIILGRNRKENLEPNAEFDKRT 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 ehltvqakapfhtppeepnecdfknmdslpsgkihrkvkiilgrnrkenlepnaefdkrt 600

Qy     601 EFITQEENRICSSPVQSLLDLFQTSEEKSEFLGFTSYTEKSGICNVLDIWEEENSDNLLT 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 efitqeenricsspvqslldlfqtseekseflgftsyteksgicnvldiweeensdnllt 660

Qy     661 AFFSSPSTSTFTGF 674
           ||||||||||||||
Db     661 affsspststftgf 674


RESULT   2
AAB03758
ID   AAB03758 standard; Protein; 674 AA.
XX
AC   AAB03758;
XX
DT   04-OCT-2000  (first entry)
XX
DE   Human H37 amino acid sequence #1.
XX
KW   H37; human; Cdc7 regulatory subunit; cytostatic; proliferative; cancer;
KW   anti-proliferative; replication regulator; stem cell.
XX
OS   Homo sapiens.
XX
PN   WO200026250-A1.
XX
PD   11-MAY-2000.
XX
PF   01-NOV-1999;   99WO-JP06076.
XX
PR   30-OCT-1998;   98JP-0311408.
XX
PA   (NISC-) JAPAN SCI & TECHNOLOGY CORP.
PA   (ARAI/) ARAI K.
PA   (MASA/) MASAI H.
XX
PI   Arai K,  Masai H;
XX
DR   WPI; 2000-365580/31.
DR   N-PSDB; AAA53483.
XX
PT   Human H37 proteins with a Cdc7 activity regulatory subunit, for
PT   controlling cell replication and cell proliferation, useful in treating
PT   cancers and diseases due to abnormal production of stem cells  -
XX
PS   Claim 1; Fig 5; 55pp; Japanese.
XX
CC   The present sequence represents a human H37 protein sequence. H37 is a
CC   protein with a Cdc7 activity regulatory subunit. The invention relates to
CC   two H37 protein and nucleotide sequences. H37 proteins exhibit
CC   cytostatic, proliferative, anti-proliferative, and cell replication
CC   regulatory activities. The proteins, encoded genes and DNA fragments are
CC   useful in treating cancers and other diseases resulting from abnormal
CC   production of stem cells. Antibodies directed against one of the H37
CC   proteins can be used to inhibit cell proliferation.
XX
SQ   Sequence   674 AA;

  Query Match            100.0%;  Score 3510;  DB 21;  Length 674;
```

```
  Best Local Similarity   100.0%;  Pred. No. 2.1e-257;
  Matches  674;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 MNSGAMRIHSKGHFQGGIQVKNEKNRPSLKSLKTDNRPEKSKCKPLWGKVFYLDLPSVTI 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 mnsgamrihskghfqggiqvkneknrpslkslktdnrpekskckplwgkvfyldlpsvti 60

Qy      61 SEKLQKDIKDLGGRVEEFLSKDISYLISNKKEAKFAQTLGRISPVPSPESAYTAETTSPH 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 seklqkdikdlggrveeflskdisylisnkkeakfaqtlgrispvpspesaytaettsph 120

Qy     121 PSHDGSSFKSPDTVCLSRGKLLVEKAIKDHDFIPSNSILSNALSWGVKILHIDDIRYYIE 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 pshdgssfkspdtvclsrgkllvekaikdhdfipsnsilsnalswgvkilhiddiryyie 180

Qy     181 QKKKELYLLKKSSTSVRDGGKRVGSGAQKTRTGRLKKPFVKVEDMSQLYRPFYLQLTNMP 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 qkkkelylikksstsvrdggkrvgsgaqktrtgrlkkpfvkvedmsqlyrpfylqltnmp 240

Qy     241 FINYSIQKPCSPFDVDKPSSMQKQTQVKLRIQTDGDKYGGTSIQLQLKEKKKKGYCECCL 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 finysiqkpcspfdvdkpssmqkqtqvklriqtdgdkyggtsiqlqlkekkkkgyceccl 300

Qy     301 QKYEDLETHLLSEQHRNFAQSNQYQVVDDIVSKLVFDFVEYEKDTPKKKRIKYSVGSLSP 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 qkyedlethllseqhrnfaqsnqyqvvddivsklvfdfveyekdtpkkkrikysvgslsp 360

Qy     361 VSASVLKKTEQKEKVELQHISQKDCQEDDTTVKEQNFLYKETQETEKKLLFISEPIPHPS 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 vsasvlkkteqkekvelqhisqkdcqeddttvkeqnflyketqetekkllfisepiphps 420

Qy     421 NELRGLNEKMSNKCSMLSTAEDDIRQNFTQLPLHKNKQECILDISEHTLSENDLEELRVD 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 nelrglnekmsnkcsmlstaeddirqnftqlplhknkqecildisehtlsendleelrvd 480

Qy     481 HYKCNIQASVHVSDFSTDNSGSQPKQKSDTVLFPAKDLKEKDLHSIFTHDSGLITINSSQ 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 hykcniqasvhvsdfstdnsgsqpkqksdtvlfpakdlkekdlhsifthdsglitinssq 540

Qy     541 EHLTVQAKAPFHTPPEEPNECDFKNMDSLPSGKIHRKVKIILGRNRKENLEPNAEFDKRT 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 ehltvqakapfhtppeepnecdfknmdslpsgkihrkvkiilgrnrkenlepnaefdkrt 600

Qy     601 EFITQEENRICSSPVQSLLDLFQTSEEKSEFLGFTSYTEKSGICNVLDIWEEENSDNLLT 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 efitqeenricsspvqslldlfqtseekseflgftsyteksgicnvldiweeensdnllt 660

Qy     661 AFFSSPSTSTFTGF 674
           ||||||||||||||
Db     661 affsspststftgf 674


RESULT   3
AAM17195
ID   AAM17195 standard; Protein; 320 AA.
XX
AC   AAM17195;
XX
DT   12-OCT-2001  (first entry)
XX
DE   Peptide #3629 encoded by probe for measuring cervical gene expression.
XX
KW   Probe; human; microarray; gene expression; cervical epithelial cell;
KW   cervical cancer.
XX
OS   Homo sapiens.
XX
PN   WO200157278-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00670.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-488901/53.
XX
PT   Human genome-derived single exon nucleic acid probes useful for
PT   analyzing gene expression in human cervical epithelial cells -
XX
PS   Claim 27; SEQ ID No 22021; 487pp; English.
XX
CC   The present invention relates to human single exon nucleic acid probes
CC   (SENP: see AAI10068-AAI28459). The present sequence is a peptide encoded
CC   by one such probe. The SENPs are derived from human HeLa cells. The SENPs
CC   can be used to produce a single exon microarray, which can be used for
CC   measuring human gene expression in a sample derived from human cervical
CC   epithelial cells. By measuring gene expression, the probes are therefore
CC   useful in grading and/or staging of diseases of the cervix, notably
CC   cervical cancer.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   320 AA;


  Query Match             47.3%;  Score 1661;  DB 22;  Length 320;
  Best Local Similarity   100.0%;  Pred. No. 7.7e-118;
  Matches  320;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     351 IKYSVGSLSPVSASVLKKTEQKEKVELQHISQKDCQEDDTTVKEQNFLYKETQETEKKLL 410
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 ikysvgslspvsasvlkkteqkekvelqhisqkdcqeddttvkeqnflyketqetekkll 60

Qy     411 FISEPIPHPSNELRGLNEKMSNKCSMLSTAEDDIRQNFTQLPLHKNKQECILDISEHTLS 470
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 fisepiphpsnelrglnekmsnkcsmlstaeddirqnftqlplhknkqecildisehtls 120

Qy     471 ENDLEELRVDHYKCNIQASVHVSDFSTDNSGSQPKQKSDTVLFPAKDLKEKDLHSIFTHD 530
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 endleelrvdhykcniqasvhvsdfstdnsgsqpkqksdtvlfpakdlkekdlhsifthd 180

Qy     531 SGLITINSSQEHLTVQAKAPFHTPPEEPNECDFKNMDSLPSGKIHRKVKIILGRNRKENL 590
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 sglitinssqehltvqakapfhtppeepnecdfknmdslpsgkihrkvkiilgrnrkenl 240

Qy     591 EPNAEFDKRTEFITQEENRICSSPVQSLLDLFQTSEEKSEFLGFTSYTEKSGICNVLDIW 650
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 epnaefdkrtefitqeenricsspvqslldlfqtseekseflgftsyteksgicnvldiw 300

Qy     651 EEENSDNLLTAFFSSPSTST 670
           ||||||||||||||||||||
Db     301 eeensdnlltaffsspstst 320


RESULT   4
AAM04892
ID   AAM04892 standard; Protein; 320 AA.
XX
AC   AAM04892;
XX
DT   09-OCT-2001  (first entry)
XX
DE   Peptide #3574 encoded by probe for measuring breast gene expression.
XX
```

```
KW   Probe; human; breast disease; breast cancer; development disorder;
KW   inflammatory disease; proliferative breast disease; non-carcinoma tumour.
XX
OS   Homo sapiens.
XX
PN   WO200157270-A2.
XX
PD   09-AUG-2001.
XX
PF   29-JAN-2001; 2001WO-US00661.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-476286/51.
XX
PT   Novel single exon nucleic acid probe used to measuring gene expression
PT   in a human breast -
XX
PS   Claim 27; SEQ ID No 13632; 322pp; English.
XX
CC   The present invention relates to novel single exon nucleic acid probes
CC   (see AAI00010-AAI10067). The present sequence is a peptide encoded by one
CC   such probe. The probes are useful for measuring human gene expression in
CC   a human breast sample, where the probe hybridises at high stringency to a
CC   nucleic acid expressed in the human breast. The probes are useful for
CC   predicting, diagnosing, grading, staging, monitoring and prognosing
CC   diseases of the human breast, particularly those diseases with polygenic
CC   aetiology. The diseases include: breast cancer, disorders of development,
CC   inflammatory diseases of the breast, fibrocystic changes, proliferative
CC   breast disease and non-carcinoma tumours.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   320 AA;

  Query Match             47.3%;  Score 1661;  DB 22;  Length 320;
  Best Local Similarity   100.0%;  Pred. No. 7.7e-118;
  Matches  320;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     351 IKYSVGSLSPVSASVLKKTEQKEKVELQHISQKDCQEDDTTVKEQNFLYKETQETEKKLL 410
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 ikysvgslspvsasvlkkteqkekvelqhisqkdcqeddttvkeqnflyketqetekkll 60

Qy     411 FISEPIPHPSNELRGLNEKMSNKCSMLSTAEDDIRQNFTQLPLHKNKQECILDISEHTLS 470
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 fisepiphpsnelrglnekmsnkcsmlstaeddirqnftqlplhknkqecildisehtls 120

Qy     471 ENDLEELRVDHYKCNIQASVHVSDFSTDNSGSQPKQKSDTVLFPAKDLKEKDLHSIFTHD 530
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 endleelrvdhykcniqasvhvsdfstdnsgsqpkqksdtvlfpakdlkekdlhsifthd 180

Qy     531 SGLITINSSQEHLTVQAKAPFHTPPEEPNECDFKNMDSLPSGKIHRKVKIILGRNRKENL 590
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 sglitinssqehltvqakapfhtppeepnecdfknmdslpsgkihrkvkiilgrnrkenl 240

Qy     591 EPNAEFDKRTEFITQEENRICSSPVQSLLDLFQTSEEKSEFLGFTSYTEKSGICNVLDIW 650
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 epnaefdkrtefitqeenricsspvqslldlfqtseekseflgftsyteksgicnvldiw 300

Qy     651 EEENSDNLLTAFFSSPSTST 670
           ||||||||||||||||||||
Db     301 eeensdnlltaffsspstst 320

RESULT   5
AAB95232
ID   AAB95232 standard; Protein; 245 AA.
XX
AC   AAB95232;
XX
DT   26-JUN-2001  (first entry)
XX
DE   Human protein sequence SEQ ID NO:17369.
XX
KW   Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX
OS   Homo sapiens.
XX
PN   EP1074617-A2.
XX
PD   07-FEB-2001.
XX
PF   28-JUL-2000; 2000EP-0116126.
XX
PR   29-JUL-1999;   99JP-0248036.
PR   27-AUG-1999;   99JP-0300253.
PR   11-JAN-2000; 2000JP-0118776.
PR   02-MAY-2000; 2000JP-0183767.
PR   09-JUN-2000; 2000JP-0241899.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T,  Isogai T,  Nishikawa T,  Hayashi K,  Saito K,  Yamamoto J;
PI   Ishii S,  Sugiyama T,  Wakamatsu A,  Nagai K,  Otsuki T;
XX
DR   WPI; 2001-318749/34.
XX
PT   Primer sets for synthesizing polynucleotides, particularly the 5602
PT   full-length cDNAs defined in the specification, and for the detection
PT   and/or diagnosis of the abnormality of the proteins encoded by the
PT   full-length cDNAs -
XX
PS   Claim 8; SEQ ID 17369; 2537pp + CD ROM; English.
XX
CC   The present invention describes primer sets for synthesising 5602
CC   full-length cDNAs defined in the specification. Where a primer set
CC   comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC   to the complementary strand of a polynucleotide which comprises one of
CC   the 5602 nucleotide sequences defined in the specification, where the
CC   oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC   of an oligonucleotide comprising a sequence complementary to the
CC   complementary strand of a polynucleotide which comprises a 5'-end
CC   sequence and an oligonucleotide comprising a sequence complementary to a
CC   polynucleotide which comprises a 3'-end sequence, where the
CC   oligonucleotide comprises at least 15 nucleotides and the combination of
CC   the 5'-end sequence/3'-end sequence is selected from those defined in
CC   the specification. The primer sets can be used in antisense therapy and
CC   in gene therapy. The primers are useful for synthesising polynucleotides,
CC   particularly full-length cDNAs. The primers are also useful for the
CC   detection and/or diagnosis of the abnormality of the proteins encoded by
CC   the full-length cDNAs. The primers allow obtaining of the full-length
CC   cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC   AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC   AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC   represent oligonucleotides, all of which are used in the exemplification
CC   of the present invention.
XX
SQ   Sequence   245 AA;

  Query Match             36.6%;  Score 1283;  DB 22;  Length 245;
  Best Local Similarity   100.0%;  Pred. No. 2.2e-89;
  Matches  245;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy    430 MSNKCSMLSTAEDDIRQNFTQLPLHKNKQECILDISEHTLSENDLEELRVDHYKCNIQAS 489
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 msnkcsmlstaeddirqnftqlplhknkqecildisehtlsendleelrvdhykcniqas 60

Qy    490 VHVSDFSTDNSGSQPKQKSDTVLFPAKDLKEKDLHSIFTHDSGLITINSSQEHLTVQAKA 549
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 vhvsdfstdnsgsqpkqksdtvlfpakdlkekdlhsifthdsglitinssqehltvqaka 120

Qy    550 PFHTPPEEPNECDFKNMDSLPSGKIHRKVKIILGRNRKENLEPNAEFDKRTEFITQEENR 609
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 pfhtppeepnecdfknmdslpsgkihrkvkiilgrnrkenlepnaefdkrtefitqeenr 180

Qy    610 ICSSPVQSLLDLFQTSEEKSEFLGFTSYTEKSGICNVLDIWEEENSDNLLTAFFSSPSTS 669
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 icsspvqslldlfqtseekseflgftsyteksgicnvldiweeensdnlltaffsspsts 240

Qy    670 TFTGF 674
          |||||
Db    241 tftgf 245


RESULT   6
AAB03759
ID   AAB03759 standard; Protein; 234 AA.
XX
AC   AAB03759;
XX
DT   04-OCT-2000  (first entry)
XX
DE   Human H37 amino acid sequence #2.
XX
KW   H37; human; Cdc7 regulatory subunit; cytostatic; proliferative; cancer;
KW   anti-proliferative; replication regulator; stem cell.
XX
OS   Homo sapiens.
XX
PN   WO200026250-A1.
XX
PD   11-MAY-2000.
XX
PF   01-NOV-1999;   99WO-JP06076.
XX
PR   30-OCT-1998;   98JP-0311408.
XX
PA   (NISC-) JAPAN SCI & TECHNOLOGY CORP.
PA   (ARAI/) ARAI K.
PA   (MASA/) MASAI H.
XX
PI   Arai K,  Masai H;
XX
DR   WPI; 2000-365580/31.
DR   N-PSDB; AAA53484.
XX
PT   Human H37 proteins with a Cdc7 activity regulatory subunit, for
PT   controlling cell replication and cell proliferation, useful in treating
PT   cancers and diseases due to abnormal production of stem cells  -
XX
PS   Claim 2; Page 46-47; 55pp; Japanese.
XX
CC   The present sequence represents a human H37 protein sequence. H37 is a
CC   protein with a Cdc7 activity regulatory subunit. The invention relates to
CC   two H37 protein and nucleotide sequences. H37 proteins exhibit
CC   cytostatic, proliferative, anti-proliferative, and cell replication
CC   regulatory activities. The proteins, encoded genes and DNA fragments are
CC   useful in treating cancers and other diseases resulting from abnormal
CC   production of stem cells. Antibodies directed against one of the H37
CC   proteins can be used to inhibit cell proliferation.
XX
SQ   Sequence   234 AA;


  Query Match           33.3%;  Score 1170;  DB 21;  Length 234;
  Best Local Similarity  100.0%;  Pred. No. 7.6e-81;
  Matches  227;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 MNSGAMRIHSKGHFQGGIQVKNEKNRPSLKSLKTDNRPEKSKCKPLWGKVFYLDLPSVTI 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 mnsgamrihskghfqggiqvkneknrpslksltkdnrpekskckplwgkvfyldlpsvti 60

Qy     61 SEKLQKDIKDLGGRVEEFLSKDISYLISNKKEAKFAQTLGRISPVPSPESAYTAETTSPH 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 seklqkdikdlggrveeflskdisylisnkkeakfaqtlgrispvpspesaytaettsph 120

Qy    121 PSHDGSSFKSPDTVCLSRGKLLVEKAIKDHDFIPSNSILSNALSWGVKILHIDDIRYYIE 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 pshdgssfkspdtvclsrgkllvekaikdhdfipsnsilsnalswgvkilhiddiryyie 180

Qy    181 QKKKELYLLKKSSTSVRDGGKRVGSGAQKTRTGRLKKPFVKVEDMSQ 227
          |||||||||||||||||||||||||||||||||||||||||||||||
Db    181 qkkkelyllkksstsvrdggkrvgsgaqktrtgrlkkpfvkvedmsq 227


RESULT   7
AAM19849
ID   AAM19849 standard; Protein; 60 AA.
XX
AC   AAM19849;
XX
DT   12-OCT-2001  (first entry)
XX
DE   Peptide #6283 encoded by probe for measuring cervical gene expression.
XX
KW   Probe; human; microarray; gene expression; cervical epithelial cell;
KW   cervical cancer.
XX
OS   Homo sapiens.
XX
PN   WO200157278-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00670.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-488901/53.
XX
PT   Human genome-derived single exon nucleic acid probes useful for
PT   analyzing gene expression in human cervical epithelial cells -
XX
PS   Claim 27; SEQ ID No 24675; 487pp; English.
XX
CC   The present invention relates to human single exon nucleic acid probes
CC   (SENP: see AAI10068-AAI28459). The present sequence is a peptide encoded
CC   by one such probe. The SENPs are derived from human HeLa cells. The SENPs
CC   can be used to produce a single exon microarray, which can be used for
CC   measuring human gene expression in a sample derived from human cervical
CC   epithelial cells. By measuring gene expression, the probes are therefore
CC   useful in grading and/or staging of diseases of the cervix, notably
CC   cervical cancer.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   60 AA;
```

```
  Query Match             8.8%;  Score 308;  DB 22;  Length 60;
  Best Local Similarity   100.0%;  Pred. No. 2.1e-16;
  Matches   60;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      74 RVEEFLSKDISYLISNKKEAKFAQTLGRISPVPSPESAYTAETTSPHPSHDGSSFKSPDT 133
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 rveeflskdisylisnkkeakfaqtlgrispvpspesaytaettsphpshdgssfkspdt 60


RESULT   8
AAM33333
ID   AAM33333 standard; Protein; 60 AA.
XX
AC   AAM33333;
XX
DT   17-OCT-2001  (first entry)
XX
DE   Peptide #7370 encoded by probe for measuring placental gene expression.
XX
KW   Probe; microarray; human; placenta; antenatal diagnosis;
KW   genetic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200157272-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00663.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-488897/53.
XX
PT   Human genome-derived single exon nucleic acid probes useful for
PT   analyzing gene expression in human placenta -
XX
PS   Claim 27; SEQ ID No 33602; 654pp; English.
XX
CC   The present invention relates to single exon nucleic acid probes (SENP:
CC   see AAI31315-AAI57546). The present sequence is a peptide encoded by one
CC   such probe. The probes are useful for producing a microarray for
CC   predicting, measuring and displaying gene expression in samples derived
CC   from human placenta. The probes are useful for antenatal diagnosis of
CC   human genetic disorders.
XX
SQ   Sequence   60 AA;


  Query Match             8.8%;  Score 308;  DB 22;  Length 60;
  Best Local Similarity   100.0%;  Pred. No. 2.1e-16;
  Matches   60;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      74 RVEEFLSKDISYLISNKKEAKFAQTLGRISPVPSPESAYTAETTSPHPSHDGSSFKSPDT 133
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 rveeflskdisylisnkkeakfaqtlgrispvpspesaytaettsphpshdgssfkspdt 60


RESULT   9
AAM32811
ID   AAM32811 standard; Protein; 55 AA.
XX
AC   AAM32811;
XX
DT   17-OCT-2001  (first entry)
XX
DE   Peptide #6848 encoded by probe for measuring placental gene expression.
XX
KW   Probe; microarray; human; placenta; antenatal diagnosis;
KW   genetic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200157272-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00663.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-488897/53.
XX
PT   Human genome-derived single exon nucleic acid probes useful for
PT   analyzing gene expression in human placenta -
XX
PS   Claim 27; SEQ ID No 33080; 654pp; English.
XX
CC   The present invention relates to single exon nucleic acid probes (SENP:
CC   see AAI31315-AAI57546). The present sequence is a peptide encoded by one
CC   such probe. The probes are useful for producing a microarray for
CC   predicting, measuring and displaying gene expression in samples derived
CC   from human placenta. The probes are useful for antenatal diagnosis of
CC   human genetic disorders.
XX
SQ   Sequence   55 AA;


  Query Match             8.0%;  Score 281;  DB 22;  Length 55;
  Best Local Similarity   100.0%;  Pred. No. 2.1e-14;
  Matches   55;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      74 RVEEFLSKDISYLISNKKEAKFAQTLGRISPVPSPESAYTAETTSPHPSHDGSSF 128
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 rveeflskdisylisnkkeakfaqtlgrispvpspesaytaettsphpshdgssf 55


RESULT  10
AAM18938
ID   AAM18938 standard; Protein; 49 AA.
XX
AC   AAM18938;
XX
DT   12-OCT-2001  (first entry)
XX
DE   Peptide #5372 encoded by probe for measuring cervical gene expression.
XX
KW   Probe; human; microarray; gene expression; cervical epithelial cell;
KW   cervical cancer.
XX
OS   Homo sapiens.
XX
PN   WO200157278-A2.
XX
PD   09-AUG-2001.
```

XX
PF 30-JAN-2001; 2001WO-US00670.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488901/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human cervical epithelial cells -
XX
PS Claim 27; SEQ ID No 23764; 487pp; English.
XX
CC The present invention relates to human single exon nucleic acid probes
CC (SENP: see AAI10068-AAI28459). The present sequence is a peptide encoded
CC by one such probe. The SENPs are derived from human HeLa cells. The SENPs
CC can be used to produce a single exon microarray, which can be used for
CC measuring human gene expression in a sample derived from human cervical
CC epithelial cells. By measuring gene expression, the probes are therefore
CC useful in grading and/or staging of diseases of the cervix, notably
CC cervical cancer.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 49 AA;

Query Match 7.4%; Score 259; DB 22; Length 49;
Best Local Similarity 100.0%; Pred. No. 8.2e-13;
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 25 NRPSLKSLKTDNRPEKSKCKPLWGKVFYLDLPSVTISEKLQKDIKDLGG 73
||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1 nrpslksiktdnrpekskckplwgkvfyldlpsvtiseklqkdikdlgg 49

RESULT 11
AAM31512
ID AAM31512 standard; Protein; 49 AA.
XX
AC AAM31512;
XX
DT 17-OCT-2001 (first entry)
XX
DE Peptide #5549 encoded by probe for measuring placental gene expression.
XX
KW Probe; microarray; human; placenta; antenatal diagnosis;
KW genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200157272-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00663.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488897/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human placenta -
XX
PS Claim 27; SEQ ID No 31781; 654pp; English.
XX
CC The present invention relates to single exon nucleic acid probes (SENP:
CC see AAI31315-AAI57546). The present sequence is a peptide encoded by one
CC such probe. The probes are useful for producing a microarray for
CC predicting, measuring and displaying gene expression in samples derived
CC from human placenta. The probes are useful for antenatal diagnosis of
CC human genetic disorders.
XX
SQ Sequence 49 AA;

Query Match 7.4%; Score 259; DB 22; Length 49;
Best Local Similarity 100.0%; Pred. No. 8.2e-13;
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 25 NRPSLKSLKTDNRPEKSKCKPLWGKVFYLDLPSVTISEKLQKDIKDLGG 73
||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1 nrpslksiktdnrpekskckplwgkvfyldlpsvtiseklqkdikdlgg 49

RESULT 12
AAB95297
ID AAB95297 standard; Protein; 170 AA.
XX
AC AAB95297;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human protein sequence SEQ ID NO:17525.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI; 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
PS Claim 8; SEQ ID 17525; 2537pp + CD ROM; English.
XX
CC The present invention describes primer sets for synthesising 5602

```
CC   full-length cDNAs defined in the specification. Where a primer set
CC   comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC   to the complementary strand of a polynucleotide which comprises one of
CC   the 5602 nucleotide sequences defined in the specification, where the
CC   oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC   of an oligonucleotide comprising a sequence complementary to the
CC   complementary strand of a polynucleotide which comprises a 5'-end
CC   sequence and an oligonucleotide comprising a sequence complementary to a
CC   polynucleotide which comprises a 3'-end sequence, where the
CC   oligonucleotide comprises at least 15 nucleotides and the combination of
CC   the 5'-end sequence/3'-end sequence is selected from those defined in
CC   the specification. The primer sets can be used in antisense therapy and
CC   in gene therapy. The primers are useful for synthesising polynucleotides,
CC   particularly full-length cDNAs. The primers are also useful for the
CC   detection and/or diagnosis of the abnormality of the proteins encoded by
CC   the full-length cDNAs. The primers allow obtaining of the full-length
CC   cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC   AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC   AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC   represent oligonucleotides, all of which are used in the exemplification
CC   of the present invention.
XX
SQ   Sequence   170 AA;


  Query Match             5.9%;  Score 208;  DB 22;  Length 170;
  Best Local Similarity   36.4%;  Pred. No. 3.5e-08;
  Matches   55;  Conservative   22;  Mismatches   44;  Indels   30;  Gaps    4;

Qy      21 KNEKNRPSLKSLKTDNRPEKSKCKPLWGKVFYLDLPSVTISEKLQKDIKDLGGRVEEFLS 80
           | :|| |  :             |  || ||||||:    : |   |: ||| :| |||
Db      35 kcqknspgark------------hpfsgksfyldlpagknlqfltgaiqqlggviegfls 82

Qy      81 KDISYLISNKKEAKFAQTLGRI-----SPVPSPESAYTAETTSPHPSHDGSSFKSPDTVC 135
           |::||::|:::| | |:: |:      || ||     |:    |  ||   | |  |:|
Db      83 kevsyivssrrevk-aessgkshrgcpspspsevrvetsamvdpkgshprpsrkpvdsvp 141

Qy     136 LSRGKLLVEKAIKDHDFIPSNSILSNALSWG 166
           ||||| |::|||:            | :|||
Db     142 lsrgkellqkair------------nqvswg 160


RESULT  13
AAB18195
ID   AAB18195 standard; Protein; 1516 AA.
XX
AC   AAB18195;
XX
DT   07-NOV-2000  (first entry)
XX
DE   Plasmodium falciparum chromosome 2 related protein SEQ ID NO:52.
XX
KW   Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
KW   antimalarial; malaria; protozoacide; infection; insecticide.
XX
OS   Plasmodium falciparum.
XX
PN   WO200025728-A2.
XX
PD   11-MAY-2000.
XX
PF   05-NOV-1999;   99WO-US26796.
XX
PR   05-NOV-1998;   98US-0107131.
XX
PA   (HOFF/) HOFFMAN S.
PA   (CARU/) CARUCCI D.
PA   (GARD/) GARDNER M.
PA   (VENT/) VENTER J C.
XX
PI   Hoffman S,  Carucci D,  Gardner M,  Venter JC;
XX
DR   WPI; 2000-365347/31.
XX
PT   Proteins encoded by chromosome 2 of the human malarial parasite,
PT   Plasmodium falciparum, useful as antimalarial vaccines and in the
PT   diagnosis of P.falciparum infection -
XX
PS   Disclosure; Page 120-124; 577pp; English.
XX
CC   The present invention describes proteins and their fragments (I) encoded
CC   by chromosome 2 of the human malarial parasite, Plasmodium falciparum.
CC   Also described are: (1) nucleotide sequences (II) encoding (I); and (2)
CC   vaccines against P. falciparum infection comprising (I) or (II).
CC   (I) and (II) are useful for the development of vaccines against
CC   P. falciparum infection. (I) and polyclonal antisera or a monoclonal
CC   antibody raised to immunogens comprising the sequences of (I), are
CC   useful in the detection of infection with P. falciparum. Furthermore,
CC   (I) (especially when they are rifins or secreted or membrane proteins)
CC   can aid the identification of drugs to treat or prevent P. falciparum
CC   infection, or they can be used to identify drug resistance in
CC   P. falciparum. Sequencing of the Plasmodium chromosome 2 and the
CC   subsequent identification of proteins encoded by it will help to expand
CC   our understanding of parasite biology, a process hampered by the
CC   complexity of the parasitic lifecycle, and provide new targets for
CC   vaccine and drug development. Parasite resistance to drugs and mosquito
CC   resistance to insecticides have led to a resurgence of malaria in many
CC   parts of the world, and there is a pressing need for vaccines and new
CC   drugs. AAA70078 to AAA70287 and AAB18144 to AAB18352 represent nucleotide
CC   and protein sequences given in the present invention, but which are not
CC   specifically mentioned within the specification.
XX
SQ   Sequence   1516 AA;


  Query Match             4.8%;  Score 170;  DB 21;  Length 1516;
  Best Local Similarity   19.5%;  Pred. No. 0.00057;
  Matches  138;  Conservative  115;  Mismatches  257;  Indels  198;  Gaps   31;

Qy      29 LKSLKTDNRPEKSKCKPLWGKVFYLDLPSVTISEKLQK------DIKDLGGRVEEFLSKD 82
           :: :|||   || | |              ::| :|      |||    :    ||
Db     646 vrmiktdfeieklkik--------------maeniqsvegellinkdlskntdninikd 690

Qy      83 ISYL---ISNKKEAKFAQTLGRISPVPSPESAYTAETTSPHPSHDGSSFKSPDTVCLSRG 139
           : |    | ||: ||   :          : |   | | :          :| : :  ::
Db     691 ynvlqkkkskkkkkkflndi---------lntynfteskyqdlyvkgeeskedi-knqi 740

Qy     140 KLLVEKAIKDHDFIPSNSILSNALSWGVKILHIDDIRYYIEQKKKELYLLKKSSTSVRDG 199
             : ::  :::| |      |:      | : ||:       || |:| |:     : :
Db     741 dfvtqecyrnndiirdthdksdif----knikidn------nkkyeiynleleqeeine- 789

Qy     200 GKRVGSGAQKTRTGRLKKPFVKVEDMSQLYRPFYLQLTNMPFINYSIQKPCSPFDVDKPS 259
                        || : |  : |   : |:|:: |
Db     790 ---------------kknynknnndsn--ktfflkien-------------------- 810

Qy     260 SMQKQTQVKLRIQTDGDKYGGTSIQLQLKEKKKKGYCECCLQK-------YEDLETHLLS 312
           :  | :  | :  | |:   :||     |    |          ||| | :
Db     811 ------efkkdllldsqifgdslladike---ynytadnldnnnenkslyedgenfit- 860

Qy     313 EQHRNFAQSNQYQVVDDIVSKLVFDFVEYE------KDTPKKKRIKYSVGSLSPVSASVL 366
             ||   :|:|:  ::|:   : |  :|       || :|:|          |    : ||
Db     861 ---rnepitneyeeknnii--yisdeqkyneediifkdkikekeknndtssddfencsvq 915

Qy     367 KKTEQKEKVELQHISQKDCQEDDTTVKEQNFLYKETQETE---KKLLFISEPIPHPSNEL 423
           :|    ||:|  :    |     :::  :   ||: :: |     |  : :   | ||:|
Db     916 ekiyvnekieeynnknddkssssssiileeikykkekkdelvspnlcvlldefeh-sndl 974

Qy     424 RGLNEKMSNKCSMLSTAEDDIRQNFTQLPLHKNKQECILDISEHTLS---ENDLEELRVD 480
                  : :| : ||:: |                :|:: ::   || :::  |:
Db     975 en---------nyisvssddmktn---------------vsknnitgvkenkvdktnve 1009

Qy     481 HYKCNIQASVHVSDFSTDNSGSQPKQKSDTVLFPAKDLKEKDLHSIFTHDSGLITINS-- 538
           : |      : :| |   :  : |   :  ::   |  ||:|   :  |     :|:
Db    1010 ydkkgddgvieis-fedshkleeskfddnnniydnddeleknlskdyisdvdknhvnniy 1068

Qy     539 -------SQEHLTVQAKAPFHTPPEEPNE--C-------------DFKNMDSLPSGKIH 575
```

```
              :|:  |: |    |   : ||  |              |  |:  | |  |:
Db   1069 niergederenefvenkiq-steshksnefictenkslrkqymskedisnvrilksddin 1127

Qy    576 RKVK-----IILGR-----NRKENLEPNAE------FDKRTEFITQEENRICSSPVQSLL 619
          |     |:| :     | : |:| | :       |:   |   |:|:|  | ::
Db   1128 nlskqnyfeilldkkqvmdnfqmnieqnndklkedkldegayfeylednkiidsyik--- 1184

Qy    620 DLFQTSEEKSEFLGFTSYTEKSGICNVLDIWEEENSD-NLLTAFFSSP 666
          :|::|  | :     :|    | ::| :| | |  ||  ||  |
Db   1185 ---etnkeneelikeykklkk----nnieindemnddiklllnffgip 1225


RESULT  14
AAB18188
ID   AAB18188 standard; Protein; 646 AA.
XX
AC   AAB18188;
XX
DT   07-NOV-2000  (first entry)
XX
DE   Plasmodium falciparum chromosome 2 related protein SEQ ID NO:45.
XX
KW   Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
KW   antimalarial; malaria; protozoacide; infection; insecticide.
XX
OS   Plasmodium falciparum.
XX
PN   WO200025728-A2.
XX
PD   11-MAY-2000.
XX
PF   05-NOV-1999;   99WO-US26796.
XX
PR   05-NOV-1998;   98US-0107131.
XX
PA   (HOFF/) HOFFMAN S.
PA   (CARU/) CARUCCI D.
PA   (GARD/) GARDNER M.
PA   (VENT/) VENTER J C.
XX
PI   Hoffman S,  Carucci D,  Gardner M,  Venter JC;
XX
DR   WPI; 2000-365347/31.
XX
PT   Proteins encoded by chromosome 2 of the human malarial parasite,
PT   Plasmodium falciparum, useful as antimalarial vaccines and in the
PT   diagnosis of P.falciparum infection -
XX
PS   Disclosure; Page 110-111; 577pp; English.
XX
CC   The present invention describes proteins and their fragments (I) encoded
CC   by chromosome 2 of the human malarial parasite, Plasmodium falciparum.
CC   Also described are: (1) nucleotide sequences (II) encoding (I); and (2)
CC   vaccines against P. falciparum infection comprising (I) or (II).
CC   (I) and (II) are useful for the development of vaccines against
CC   P. falciparum infection. (I) and polyclonal antisera or a monoclonal
CC   antibody raised to immunogens comprising the sequences of (I), are
CC   useful in the detection of infection with P. falciparum. Furthermore,
CC   (I) (especially when they are rifins or secreted or membrane proteins)
CC   can aid the identification of drugs to treat or prevent P. falciparum
CC   infection, or they can be used to identify drug resistance in
CC   P. falciparum. Sequencing of the Plasmodium chromosome 2 and the
CC   subsequent identification of proteins encoded by it will help to expand
CC   our understanding of parasite biology, a process hampered by the
CC   complexity of the parasitic lifecycle, and provide new targets for
CC   vaccine and drug development. Parasite resistance to drugs and mosquito
CC   resistance to insecticides have led to a resurgence of malaria in many
CC   parts of the world, and there is a pressing need for vaccines and new
CC   drugs. AAA70078 to AAA70287 and AAB18144 to AAB18352 represent nucleotide
CC   and protein sequences given in the present invention, but which are not
CC   specifically mentioned within the specification.
XX
SQ   Sequence   646 AA;
```

```
  Query Match             4.7%;  Score 164;  DB 21;  Length 646;
  Best Local Similarity   18.5%;  Pred. No. 0.00049;
  Matches  136;  Conservative  116;  Mismatches  229;  Indels  256;  Gaps   28;

Qy      39 EKSKCKPLWGKVFYLDLPSVTISEKLQKDIKDLGGRVEEFLSKDISYLISNKKEAKFAQT 98
           |::|   |  | | ::   ||   |  ||| :|  |:|    ::|  |||:  :  | :
Db      10 eqnkmntlhiknfimenlkvteeikhdkdinnlmrrieh---eeikelissngkryfme- 65

Qy      99 LGRISPVPSPESAYTAETTSPHPSHD---GSSFKSPDTV------CLSRGKLLVEKAIKD 149
             |  :      :  |   |  : |     |||    :       : :  : ::| :|
Db      66 ---irkiyflmkkfhkegyfpssnkdvlkkqsfkrnkniknllqesikkkniqiqkllkq 122

Qy     150 H------------DFIPSNSILSNALSWG------------------------------ 166
           :            :   | :| :  |:
Db     123 yiilkgyyknvckkyrqenellksffsfsnnqsyylnlkyspphsrrnriyfypytkllr 182

Qy     167 ----VKILHIDDIRYYI------EQKKKELYLLKK------------------------ 191
               :| |  : || |      ::|||::|: ||
Db     183 rkrlrrishfkedryvihkgpltkkkkkkiyinkkyiyiiyiyiyiyyiffmfysfifie 242

Qy     192 -SSTSVRDGGKRVGSGAQKTRTGRLKKPFVKVEDMSQLYRPFYL------QLTNMPFINY 244
            | |:                          | |      : |:           :| |: |  |
Db     243 yfsnsi---------------------frkythhkkrykeiiqdilndnkllnlhfkry 280

Qy     245 SIQKPCSPFDVDKPSSMQKQTQVKLRIQTDGDKYGGTSIQLQLKEKKKKGYCECCLQKYE 304
                       :|: : || | :   |    :: |  |:|||           |
Db     281 k------------ekykkkkkkklhisskrkk-dkrnldlyckkkkk----------e 315

Qy     305 DLETHLL--SEQHRNFAQSNQYQVV----DDIVSKLVFDFVEYEKDTPKKKRIKYSVGSL 358
             : |||   :       :|: |  :    ::|::|   : :  |::
Db     316 iiythlflptrlrekinkssnynylnkegeniinkeeenilhkeee------------- 361

Qy     359 SPVSASVLKKTEQK-EKVELQHISQKDCQEDDTTVKEQNFLYKETQETEKKLLFISEPIP 417
                :| | |:   | | ::|  || :|:    :|:| |:|: :|
Db     362 -----hilhkdeenymkeeeenilhkdeeenilykeeenilhkdeee------------- 403

Qy     418 HPSNELRGLNEKMSNKCSMLSTAEDDIRQNFTQLPLHKNKQECILDISEHTLSENDLEEL 477
                          ::|   |::|        |||:::| ||   |  :   | ||
Db     404 -----------------nilykeeeni--------lhkdeeenilhkeeenilhkdeeen 438

Qy     478 RVDHYKCNIQASVHVSDFSTDNSGSQPKQKSDTVLFPAKDLKEKDLHSIFTHDSGLITIN 537
             :   : ||      :   | |  :: |:| :|:     :  |:|:  :         :| |
Db     439 ilykeeenilhkeeaniietkn--aevkkkntl----rkkkkkekknflndhmkevtkn 492

Qy     538 SSQEHLTVQAKAPFHTPPEEPNECDFKNMDSLPSGKIHRKVKIILGRNRKENL------- 590
             :              :: :: |  ||      |:  | |     : |||:
Db     493 dddd-------------ddddddddennmi-----kveekqk-yndedgkenvsidnvee 532

Qy     591 --EPNAEFDKRTEFIT--QEENRICSSPVQSLLDLFQTSEEK-----SEFLGFTSYTEKS 641
               :  |:||:   ::  :||| |  |  |::  :  |::||      :    :   :|:
Db     533 cnkmkdeydkkennvsnieeeniildskeqni--ildtnkekliskekkkkkisrkikkt 590

Qy     642 GICNVLDIWEEENSDNL 658
             | :  || | || : :
Db     591 kiednkdikenenfnei 607


RESULT  15
AAY19935
ID   AAY19935 standard; Protein; 1087 AA.
XX
AC   AAY19935;
XX
DT   19-JUL-1999  (first entry)
XX
DE   B. burgdorferi antigenic protein, t742.aa.
XX
KW   Antigenic protein; vaccine; Lyme disease; infection; detection.
XX
OS   Borrelia burgdorferi.
XX
```

PN WO9859071-A1.
XX
PD 30-DEC-1998.
XX
PF 18-JUN-1998; 98WO-US12718.
XX
PR 03-SEP-1997; 97US-0057483.
PR 20-JUN-1997; 97US-0050359.
PR 22-JUL-1997; 97US-0053344.
PR 22-JUL-1997; 97US-0053377.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PA (MEDI-) MEDIMMUNE INC.
XX
PI Choi GH, Erwin AL, Hanson MS, Lathigra R;
XX
DR WPI; 1999-189980/16.
DR N-PSDB; AAX61632.
XX
PT New isolated Borrelia burgdorferi nucleic acids - used to develop
PT products for the diagnosis, prevention and treatment of diseases
PT caused by Borrelia, particularly Lyme disease
XX
PS Claim 12; Page 125; 275pp; English.
XX
CC This sequence represents a Borrelia burgdorferi (Bb) protein of the
CC invention, which is suitable for use in a vaccine. The Bb polypeptides
CC can be used in vaccines for eliciting protective antibodies to members of
CC the Borrelia genus, particularly for the use against Lyme disease in
CC humans and animals. They can be used for preventing or attenuating an
CC infection caused by a member of the Borrelia genus. The products can also
CC be used for detection of members of the Borrelia genus.
XX
SQ Sequence 1087 AA;

Query Match 4.7%; Score 164; DB 20; Length 1087;
Best Local Similarity 19.2%; Pred. No. 0.001;
Matches 157; Conservative 126; Mismatches 265; Indels 270; Gaps 42;

```
Qy      22 NEKNRPSLKSLKTDNRPEKSKCKP---LWGKVF--YLDLPSVTISEKLQKDIKDLG---- 72
           |  | ||| : :::: |    |   : ||::  |    |  | ::| : : |:
Db     217 nnnnttslkkissnsqkeselsppsqtiigkiyrpy----sylikkelyeilddintgrv 272

Qy      73 ----GRVEEFLSKDIS-------YLISNKKEAKFAQTL-----GRISP--VPSPESAY-- 112
                |::| : | :|        || | |  : :  |      | |  :  |:  |
Db     273 tlgknrlkelikkglsnkfqkvnelienskneasnllltlikkdiepnlinipkdpykk 332

Qy     113 -----TAETTSPHPSHD-GSSFKSPDTVCLSRGKLLVEKAIKD-HDFI---PSNSILSNA 162
                |  |   |  |  |   : |   |   ::|||| ::|:   |:::  |
Db     333 eifqldkedkkpqyledlkskvhsikpidlentksr-qqaikdlneflknnpndaqaskt 391

Qy     163 LSWGVKILHIDDIRYYIEQ-KKKELYLLKKSSTSVRDGG---KRVGSGAQKTRTGRLKKP 218
           |:   || |::|::  :   |  :|   |    :::|     |   : || ::|
Db     392 laqankiqhledlkskvhsikpidlentksrqqaikdlneflknnpndaqasktlaqank 451

Qy     219 FVKVEDMSQLYRPFYLQLTNMPFINYSIQKPCSPFDVDKPSSMQKQTQVKLRIQTDGDKY 278
              :||:              :||:   | |::   | |
Db     452 iqhledlkskv-------------hsik----pidlentksrq---------------- 477

Qy     279 GGTSIQLQLKEKKKKGYCECCLQKYEDLETHLL-----SEQHRNFAQSNQYQVVDDIVSK 333
                                 |  :||   |      ::  :  ||:|: | ::|: ||
Db     478 ---------------------qaikdlneflknnpndaqasktlaqankiqhledlksk 515

Qy     334 L-----------------VFDFVEYEKDTPKKKR----------------IKYSVGSLSP 360
           :               : |  |: |: | :                 :|  | |: |
Db     516 vhsikpidlentksrqqaikdlnefxknnpndaqasktlaqankiqhledlkskvhsikp 575

Qy     361 VSASVLKKTEQ--KEKVELQHISQKDCQEDDTTVKEQNFLYKETQETEKKLLFISEPIPH 418
           :     |  :|  |:  |    :  | |   |  :     : |  : : |:  | :||
Db     576 idlentksrqqaikdlnefxknnpndaqasktlaqankiqhle--dlkskvhsi-kpidl 632

Qy     419 PSNE-----LRGLNEKMSNKCSMLSTAEDDIRQNFTQ----------------LPLHKNK 457
                                                                  |::
Db     633 entksrqqaikdlneflknnpndaqasktlaqankiqhledlkskvhsikpidlentksr 692

Qy     458 QECILDISE------------HTL------------SENDLEEL------RVDHYKCNI- 486
           |: | |::|             ||            :||  |::      : ||||  |
Db     693 qqaikdlnefxknnpndaqasktlaqayenngdllkaenayekiikltntqedhyklgii 752

Qy     487 -------QASVHVSDFSTDNSGSQPKQKSDTVLFPAKDLKEKDLHSIFTHDSG--LITIN 537
                  : |:   |                          : :|    |    |: |  |: :|
Db     753 rfklkkyehsiesfd--------------------qtikldpkhkkalhnkgialmmln 791

Qy     538 SSQEHL-----TVQAKAPFHTP------PEEPNE------CDFK---NMDSLPSGKIHRK 577
             ::: :      :|    : |         || |          ||   |:|  |: :  |
Db     792 knkkaiesfekaiqidknygtayyqkgiaeeknqdmqqafasfknaynldknpnyal--k 849

Qy     578 VKII---LGRNRKENLEP----NAEFDKRTEF------ITQEENRICSSPVQSLLDLFQT 624
             |:   || | |:: |      ||    |  |       | : ||      ::::
Db     850 agivsnnlg-nfkqseeylnffnanakkpneiaiynlsiakfennkleesletinkaidl 908

Qy     625 SEEKSEFLGFTS--YTEKSGICNVLDIWE---EENSDN 657
           : ||||:|   :    :|    | : ::    |:| :|
Db     909 npekseylylkasinlkkenyqnaislyslvieknpen 946
```

Search completed: December 27, 2001, 16:53:19
Job time: 519 sec